PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 9/64, C12Q 1/68, C12N 15/57

A1

(11) International Publication Number: **WO 99/33963**

(43) International Publication Date: 8 July 1999 (08.07.99)

(21) International Application Number: PCT/US98/26547

(22) International Filing Date: 14 December 1998 (14.12.98)

(30) Priority Data:
60/070,112 31 December 1997 (31.12.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/070,112 (CIP)
Filed on 31 December 1997 (31.12.97)

(71) Applicant *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street – R440, Emeryville, CA 94608 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GIESE, Klaus, W. [US/US]; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94662–8097 (US). XIN, Hong [US/US]; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94662–8097 (US).

(74) Agents: POTTER, Jane, E., R. et al.; Chiron Corporation, Intellectual Property – R440, P.O. Box 8097, Emeryville, CA 94662–8097 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METASTATIC CANCER REGULATED GENE

(57) Abstract

A novel aspartyl protease gene termed *CSP56* can be used to provide reagents and methods for determining which tumors are likely to metastasize and for suppressing metastases of these tumors. Clinicians can use this information to predict which tumors will metastasize to other organs and to provide relevant therapies to appropriate patients.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

METASTATIC CANCER REGULATED GENE

TECHNICAL FIELD OF THE INVENTION

This invention relates to methods for predicting the behavior of tumors. In particular, the invention relates to methods in which a tumor sample is examined for expression of a specified gene sequence in order to determine its propensity for metastatic spread.

BACKGROUND OF THE INVENTION

The pathogenesis of cancer metastasis, such as breast cancer metastasis, consists of a series of linked, and selective steps including invasion, detachment, intravasion, circulation, adhesion, extravasion, and growth in distant organs (Fidler & Radinsky, *J. Natl. Cancer Inst. 88*, 1700-03, 1995). Invasiveness, one of the initial steps in metastasis, requires the expression of degradative enzymes such as plasminogen activator (Sappino *et al.*, *Cancer Res. 47*, 4043-46, 1989), collagenase (Ogilvie *et al.*, *J. Natl. Cancer Inst. 74*, 19-27, 1985), and cathepsins (Rochefort *et al.*, *J. Cell. Biochem. 35*, 17-29, 1987).

Despite the use of a number of histochemical, genetic, and immunological markers, clinicians still have a difficult time predicting which tumors will metastasize to other organs. Some patients are in need of adjuvant therapy to prevent recurrence and metastasis and others are not. However, distinguishing between these subpopulations of patients is not straightforward, and the course of treatment is not easily charted. Thus, there is a need in the art for new markers for determining which tumors are likely to metastasize.

SUMMARY OF THE INVENTION

It is an object of the invention to provide reagents and methods for determining which tumors are likely to metastasize and for suppressing metastases of these tumors. These and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention is an isolated human CSP56 protein having an amino acid sequence which is at least 85% identical to SEQ ID NO:2. Percent identity between the first and second human CSP56 proteins is determined using a Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

Another embodiment of the invention is an isolated polypeptide comprising at least 8 contiguous amino acids as shown in SEQ ID NO:2.

Even another embodiment of the invention is a CSP56 fusion protein comprising a first protein segment and a second protein segment fused together by means of a peptide bond. The first protein segment consists of at least 8 contiguous amino acids of a human CSP56 protein having an amino acid sequence as shown in SEQ ID NO:2.

Yet another embodiment of the invention is a preparation of antibodies which specifically bind to a human CSP56 protein having an amino acid sequence as shown in SEQ ID NO:2.

Still another embodiment of the invention is a cDNA molecule which encodes a human CSP56 protein having an amino acid sequence which is at least 85% identical to SEQ ID NO:2. Percent identity is determined using a Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

Even another embodiment of the invention is a cDNA molecule which encodes at least 8 contiguous amino acids of SEQ ID NO:2.

Another embodiment of the invention is a cDNA molecule which comprises at least 12 contiguous nucleotides of SEQ ID NO:1.

Yet another embodiment of the invention is a cDNA molecule which is at least 85% identical to the nucleotide sequence shown in SEQ ID NO:1. Percent identity is

determined using a Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

Still another embodiment of the invention is an isolated and purified subgenomic polynucleotide comprising a nucleotide sequence which hybridizes to SEQ ID NO:1 after washing with 0.2X SSC at 65 °C. The nucleotide sequence encodes a CSP56 protein having the amino acid sequence of SEQ ID NO:2.

Even another embodiment of the invention is a construct comprising a promoter and a polynucleotide segment encoding at least 8 contiguous amino acids of a human CSP56 protein as shown in SEQ ID NO:2. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter.

Another embodiment of the invention is a host cell comprising a construct. The construct comprises a promoter and a polynucleotide segment which encodes at least 8 contiguous amino acids of a human CSP56 protein having an amino acid sequence as shown in SEQ ID NO:2.

Yet another embodiment of the invention is a recombinant host cell comprising a new transcription initiation unit. The new transcription initiation unit comprises in 5' to 3' order: an exogenous regulatory sequence, an exogenous exon, and a splice donor site. The new transcription initiation unit is located upstream of a coding sequence of an *CSP56* gene having a coding sequence as shown in SEQ ID NO:1. The exogenous regulatory sequence controls transcription of the coding sequence of the *CSP56* gene.

Still another embodiment of the invention is a polynucleotide probe comprising at least 12 contiguous nucleotides of SEQ ID NO:1.

Yet another embodiment of the invention is a method of diagnosing neoplasia. An expression product of the nucleotide sequence shown in SEQ ID NO:1 is detected in a body sample. Detection of the expression product identifies the body sample as neoplastic.

Still another embodiment of the invention is a method for determining metastatic potential of a tumor. An expression product of a gene having the coding sequence shown in SEQ ID NO:1 is measured in a tumor sample. A tumor sample which expresses the

expression product is categorized as having metastatic potential.

Even another embodiment of the invention is a method of screening test compounds for the ability to suppress the metastatic potential of a tumor. A cell is contacted with a test compound. Synthesis of a protein having the amino acid sequence shown in SEQ ID NO:2 is measured in the cell. A test compound which decreases the amount of the protein synthesized in the cell is identified as a potential agent for suppressing the metastatic potential of the tumor.

Yet another embodiment of the invention is a set of primers for amplifying at least a portion of a gene having the coding sequence shown in SEQ ID NO:1.

The invention thus provides the art with a means of diagnosing, prognosing, and treating tumors with high metastatic potential.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Arbitrary primer-based differential display and confirmation by RNA blot analysis of different human breast cancer cell line. Figure 1A. Autoradiograph of a differential display gel depicting two bands of approximately 1.2 kb in size in the human breast cancer cell line MDA-MB-435. Differential display reactions were prepared and run in duplicates. Figure B. Northern blot analysis verifying the expression pattern in MDA-MB-435. cDNA isolated from the differential display gel hybridized to two transcripts of approximately 2.0 kb and 2.5 kb in size. Equal amounts of RNA in each lane were loaded as judged by staining of the membrane with methylene blue and hybridization of the membrane with a human β-actin probe.

Figure 2. Nucleotide sequence and deduced amino acid sequence of CSP56. Figure 2A. The 518 amino acid long sequence is shown in single-letter code below the nucleotide sequence of 1855 base pairs. The active site residue (D) and flanking amino acid residues characteristic of aspartyl proteases are underlined. The putative propeptide is boxed. The putative signal peptide at the N-terminus and the transmembrane domain at the C-terminus are underlined. Figure 2B. Expressed sequence tags extending the nucleotide sequence of CSP56 to 2606 base pairs in length. Figure 2C. Schematic representation of CSP56. SS, signal sequence; Pro, propeptide; TM transmembrane

domain. The asterisks indicate the active sites.

Figure 3. Multiple amino acid sequence alignment of CSP56 with other members of the pepsin family of aspartyl proteases. Identical amino acid residues are indicated by black boxes. The aspartyl protease active residues (D-S/T-G) are indicated by a bar on top. The cysteine residues characteristic for aspartyl protease in members of the pepsin family are indicted by asterisks. The putative membrane attachment domain is underlined. Gaps are indicated by dots. Cat-E, cathepsin E; Pep-A, pepsinogen E; Pep-C, pepsinogen C; Cat-D, cathepsin D.

Figure 4. CSP56 expression in primary tumor and metastases isolated from scid mice. Northern blot analysis using RNA isolated from primary tumors (PT) and metastatic tissues (Met) of mice injected with different human breast cancer cell lines. Equal amounts of RNA in each lane were loaded as judged by staining of the membrane with methylene blue and hybridization of the membrane with a human β-action probe.

Figure 5. CSP56 is up-regulated in patient breast tumor samples. Figure 5A. Northern blot analysis using RNA isolated from tumor and normal breast tissue from the same patient. Figure 5B. Northern blot analysis using RNA isolated from three different human breast tumor patients and normal breast tissue.

Figure 6. *In situ* hybridization analysis of CSP56 expression in breast and colon tumors. Adjacent or near-adjacent sections through normal breast tissue (A-C) and the primary breast tissue (D-F) of one patient and through normal colon tissue (G, H), the primary colon tumor (J, K), and the liver metastatis (L, M) of another patient. Sections A, D, G, J, and L were stained with haematoxylin and eosin (H & E). Sections B, E, H, K, and M were hybridized with the antisense CSP56 probe, and sections C and F were hybridized with the CSP56 sense control probe. d, lactiferous duct; f, fatty connective tissue; ly, lymphocytes; m, colon mucosa; met, metastatic tissue; PT, primary tumor; st, stroma; tc, tumor cells.

Figure 7. Expression of CSP56 in human tissues. RNA blot analysis depicting two CSP56 transcripts of 2.0 kb and 2.5 kb in various human tissues. sk. muscle, skeletal muscle; sm. intestine, small intestine; p.b. lymphocytes, peripheral blood lymphocytes.

DETAILED DESCRIPTION

It is a discovery of the present invention that a novel aspartyl-type protease, CSP56, is over-expressed in highly metastatic cancer, particularly in breast and colon cancer, and is associated with the progression of primary tumors to a metastatic state. This information can be utilized to make diagnostic reagents specific for expression products of the *CSP56* gene. It can also be used in diagnostic and prognostic methods which will help clinicians to plan appropriate treatment regimes for cancers, especially of the breast and colon.

The amino acid sequence of CSP56 protein is shown in SEQ ID NO:2. Either the CSP56 protein shown in SEQ ID NO:2 or naturally or non-naturally occurring biologically active protein variants of CSP56 protein can be used in diagnostic and therapeutic methods of the invention. Biologically active CSP56 variants retain the same biological activities as the CSP56 protein shown in SEQ ID NO:2. Biological activities of CSP56 proteins include differential expression between tumors and normal tissue, particularly between tumors with high metastatic potential and normal tissue, the ability to permit metastases, and aspartyl-type protease activity.

Biological activity of a CSP56 variant can be readily determined by one of skill in the art. Differential expression of the variant, for example, can be measured in cell lines which vary in metastatic potential, such as the breast cancer cell lines MDA-MB-231 (Brinkley *et al.*, *Cancer Res. 40*, 3118-29, 1980), MDA-MB-435 (Brinkley *et al.*, 1980), MCF-7, BT-20, ZR-75-1, MDA-MB-157, MDA-MB-361, MDA-MB-453, Alab and MDA-MB-468, or colon cancer cell lines Km12C and Km12L4A. The MDA-MB-231 cell line was deposited at the ATCC on May 15, 1998 (ATCC CRL-12532). The Km12C cell line was deposited at the ATCC on May 15, 1998 (ATCC CRL-12533). The Km12L4A cell line was deposited at the ATCC on March 19, 1998 (ATCC CRL-12496). The MDA-MB-435 cell line was deposited at the ATCC on October 9, 1998 (ATCC CRL 12583). The MCF-7 cell line was deposited at the ATCC on October 9, 1998 (ATCC CRL-12584).

Expression in a non-cancerous cell line, such as the breast cell line Hs58Bst, can

be compared with expression in cancerous cell lines. Alternatively, a breast cancer cell line with high metastatic potential, such as MDA-MB-231 or MDA-MB-435, can be contacted with a polynucleotide encoding a variant and assayed for lowered metastatic potential, for example by monitoring cell division or protein or DNA synthesis, as is known in the art. Aspartyl protease activity of a potential variant can also be measured, for example, as taught in Wright *et al.*, *J. Prot. Chem. 16*, 171-81 (1997).

Naturally occurring biologically active CSP56 protein variants are found in humans or other species and comprise amino acid sequences which are substantially identical to the amino acid sequence shown in SEQ ID NO:2. Non-naturally occurring biologically active CSP56 protein variants can be constructed in the laboratory, using standard recombinant DNA techniques. Preferably, naturally or non-naturally occurring biologically active CSP56 protein variants have amino acid sequences which are at least 65%, 75%, 85%, 90%, or 95% identical to the amino acid sequence shown in SEQ ID NO:2 and have similar differential expression patterns and aspartyl-type protease activity, though these properties may differ in degree. More preferably, the variants are at least 98% or 99% identical. Percent sequence identity between the protein of SEQ ID NO:2 and a biologically active variant can be determined using computer programs which employ the Smith-Waterman algorithm using an affine gap search with the following parameters: a gap open penalty of 12 and a gap extension penalty of 1. The Smith-Waterman homology search algorithm is taught in Smith and Waterman, *Adv. Appl. Math.* (1981) *2*:482-489.

Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, such as DNASTAR software. Preferably, amino acid changes in biologically active CSP56 protein variants are conservative amino acid changes, *i.e.*, substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine,

methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. It is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the biological properties of the resulting CSP56 protein variant, especially if the replacement is not at the catalytic domains of the protease.

CSP56 protein variants also include allelic variants, species variants, muteins, glycosylated forms, aggregative conjugates with other molecules, and covalent conjugates with unrelated chemical moieties which retain biological activity. Truncations or deletions of regions which do not affect the expression patterns or aspartyl protease activity of CSP56 protein are also biologically active CSP56 variants. Covalent CSP56 variants can be prepared by linkage of functionalities to groups which are found in the amino acid chain or at the N- or C-terminal residue, as is known in the art.

A subset of mutants, called muteins, is a group of polypeptides with the non-disulfide bond participating cysteines substituted with a neutral amino acid, generally, with serines. These mutants may be stable over a broader temperature range than kismet. See Mark *et al.*, U.S. Pat. No. 4,959,314.

CSP56 polypeptides contain less than full-length CSP56. For example, CSP56 polypeptides can contain at least 8, 10, 11, 12, 13, 14, 15, 16, 20, 21, 23, 25, 28, 29, 30, 31, 33, 35, 40, 50, 60, 75, 100, or 112 or more amino acids of a CSP56 protein or biologically active variant in the same order as found in a CSP56 protein or biologically active variant. As described above for CSP56 protein variants, polypeptide molecules having substantially the same amino acid sequence as the amino acid sequence shown in SEQ ID NO:2 but possessing minor amino acid substitutions which do not substantially affect the biological properties of a particular CSP56 polypeptide variant are within the definition of CSP56 polypeptides. Preferred CSP56 polypeptides comprise at least amino acids 106-115, 105-116, 104-117, 100-120, 297-306, 296-307, 295-308, 290-320,

8-20, 7-21, 6-22, 1-30, 461-489, 460-490, 459-491, and 407-518 of SEQ ID NO:2.

CSP56 protein or polypeptides can be isolated from, for example, MDA-MB-435 cells, using biochemical techniques well known to the skilled artisan. A preparation of isolated and purified CSP56 protein is at least 80% pure; preferably, the preparations are at least 90%, 95%, 98%, or 99% pure. CSP56 proteins and polypeptides can also be produced by recombinant DNA methods or by synthetic chemical methods. For production of recombinant CSP56 proteins or polypeptides, coding sequences selected from the *CSP56* nucleotide sequence shown in SEQ ID NO:1 can be expressed in known prokaryotic or eukaryotic expression systems. Bacterial, yeast, insect, or mammalian expression systems can be used, as is known in the art. Alternatively, synthetic chemical methods, such as solid phase peptide synthesis, can be used to synthesize CSP56 protein or polypeptides. Biologically active CSP56 protein or polypeptide variants can be similarly produced.

Fusion proteins comprising at least 8, 10, 11, 12, 13, 14, 15, 16, 20, 21, 23, 25, 28, 29, 30, 31, 33, 35, 40, 50, 60, 75, 100, or 112 contiguous CSP56 amino acids can also be constructed. CSP56 fusion proteins are useful for generating antibodies against CSP56 amino acid sequences and for use in various assay systems. For example, CSP56 fusion proteins can be used to identify proteins which interact with CSP56 protein and influence, for example, its aspartyl protease activity, its differential expression, or its ability to permit metastases. Physical methods, such as protein affinity chromatography, or library-based assays for protein-protein interactions, such as the yeast two-hybrid or phage display systems, can also be used for this purpose. Such methods are well known in the art and can also be used as drug screens.

A CSP56 fusion protein comprises two protein segments fused together by means of a peptide bond. The first protein segment consists of at least 8, 10, 11, 12, 13, 14, 15, 16, 20, 21, 23, 25, 28, 29, 30, 31, 33, 35, 40, 50, 60, 75, 100, or 112 contiguous amino acids of a CSP56 protein. The amino acids can be selected from the amino acid sequence shown in SEQ ID NO:2 or from a biologically active variant of that sequence, such as those described above. The first protein segment can also be a full-length CSP56 protein. The first protein segment can be N-terminal or C-terminal, as is convenient.

The second protein segment can be a full-length protein or a protein fragment or polypeptide. Proteins commonly used in fusion protein construction include β-galactosidase, β-glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). Additionally, epitope tags are used in fusion protein constructions, including histidine (His) tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Other fusion constructions can include maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

These fusions can be made, for example, by covalently linking two protein segments or by standard procedures in the art of molecular biology. Recombinant DNA methods can be used to prepare CSP56 fusion proteins, for example, by making a DNA construct which comprises coding sequences selected from SEQ ID NO:1 in proper reading frame with nucleotides encoding the second protein segment and expressing the DNA construct in a host cell, as is known in the art. Many kits for constructing fusion proteins are available from companies that supply research labs with tools for experiments, including, for example, Promega Corporation (Madison, WI), Stratagene (La Jolla, CA), Clontech (Mountain View, CA), Santa Cruz Biotechnology (Santa Cruz, CA), MBL International Corporation (MIC; Watertown, MA), and Quantum Biotechnologies (Montreal, Canada; 1-888-DNA-KITS).

Isolated CSP56 proteins, polypeptides, biologically active variants, or fusion proteins can be used as immunogens, to obtain a preparation of antibodies which specifically bind to epitopes of CSP56 protein. The antibodies can be used, *inter alia*, to detect CSP56 protein in human tissue, particularly in human tumors, or in fractions thereof. The antibodies can also be used to detect the presence of mutations in the *CSP56* gene which result in under- or over-expression of the CSP56 protein or in expression of a CSP56 protein with altered size or electrophoretic mobility. By binding to CSP56, antibodies can also prevent CSP56 aspartyl-type protease activity or the ability of CSP56 to permit metastases.

Antibodies which specifically bind to epitopes of CSP56 proteins, polypeptides, fusion proteins, or biologically active variants can be used in immunochemical assays, including but not limited to Western Blots, ELISAs, radioimmunoassays, immunohistochemical assays, immunoprecipitations, or other immunochemical assays known in the art. Typically, antibodies of the invention provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in such immunochemical assays. Preferably, antibodies which specifically bind to CSP56 epitopes do not detect other proteins in immunochemical assays and can immunoprecipitate CSP56 protein or polypeptides from solution.

CSP56-specific antibodies specifically bind to epitopes present in a CSP56 protein having the amino acid sequence shown in SEQ ID NO:2 or to biologically active variants of that sequence. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. However, epitopes which involve non-contiguous amino acids may require more, *e.g.*, at least 15, 25, or 50 amino acids. Preferably, CSP56 epitopes are not present in other human proteins, particularly in other aspartyl proteases.

Epitopes of CSP56 which are particularly antigenic can be selected, for example, by routine screening of CSP56 polypeptides for antigenicity or by applying a theoretical method for selecting antigenic regions of a protein to the amino acid sequence shown in SEQ ID NO:2. Such methods are taught, for example, in Hopp and Wood, *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824-28 (1981), Hopp and Wood, *Mol. Immunol. 20*, 483-89 (1983), and Sutcliffe *et al.*, *Science 219*, 660-66 (1983). By reference to Figure 3, antigenic regions of CSP56 which could also bind to antibodies which bind to other aspartyl proteases can be avoided.

Any type of antibody known in the art can be generated to bind specifically to CSP56 epitopes. For example, preparations of polyclonal and monoclonal antibodies can be made using standard methods which are well known in the art. Similarly, single-chain antibodies can also be prepared. Single-chain antibodies which specifically bind to CSP56 epitopes can be isolated, for example, from single-chain immunoglobulin display libraries, as is known in the art. The library is "panned" against CSP56 amino acid sequences, and a number of single chain antibodies which bind with high-affinity to

different epitopes of CSP56 protein can be isolated. Hayashi *et al.*, 1995, *Gene 160*:129-30. Single-chain antibodies can also be constructed using a DNA amplification method, such as the polymerase chain reaction (PCR), using hybridoma cDNA as a template. Thirion *et al.*, 1996, *Eur. J. Cancer Prev. 5*:507-11.

Single-chain antibodies can be mono- or bispecific, and can be bivalent or tetravalent. Construction of tetravalent, bispecific single-chain antibodies is taught, for example, in Coloma and Morrison, 1997, *Nat. Biotechnol. 15*:159-63. Construction of bivalent, bispecific single-chain antibodies is taught *inter alia* in Mallender and Voss, 1994, *J. Biol. Chem. 269*:199-206.

A nucleotide sequence encoding a single-chain antibody can be constructed using manual or automated nucleotide synthesis, cloned into an expression construct using standard recombinant DNA methods, and introduced into a cell to express the coding sequence, as described below. Alternatively, single-chain antibodies can be produced directly using, for example, filamentous phage technology. Verhaar *et al.*, 1995, *Int. J. Cancer 61*:497-501; Nicholls *et al.*, 1993, *J. Immunol. Meth. 165*:81-91.

Monoclonal and other antibodies can also be "humanized" in order to prevent a patient from mounting an immune response against the antibody when it is used therapeutically. Such antibodies may be sufficiently similar in sequence to human antibodies to be used directly in therapy or may require alteration of a few key residues. Sequence differences between, for example, rodent antibodies and human sequences can be minimized by replacing residues which differ from those in the human sequences, for example, by site directed mutagenesis of individual residues, or by grating of entire complementarity determining regions. Alternatively, one can produce humanized antibodies using recombinant methods, as described in GB2188638B. Antibodies which specifically bind to CSP56 epitopes can contain antigen binding sites which are either partially or fully humanized, as disclosed in U.S. 5,565,332.

Other types of antibodies can be constructed and used therapeutically in methods of the invention. For example, chimeric antibodies can be constructed as disclosed, for example, in WO 93/03151. Binding proteins which are derived from immunoglobulins and which are multivalent and multispecific, such as the "diabodies" described in WO

94/13804, can also be prepared.

Antibodies of the invention can be purified by methods well known in the art. For example, antibodies can be affinity purified by passing the antibodies over a column to which an CSP56 protein, polypeptide, variant, or fusion protein is bound. The bound antibodies can then be eluted from the column, using a buffer with a high salt concentration.

The invention also provides isolated polynucleotides which encode CSP56 protein, polypeptides, variants, or fusion proteins. Isolated polynucleotides contain less than a whole chromosome. Preferably, the polynucleotides are intron-free. An isolated *CSP56* polynucleotide encodes at least 8, 10, 12, 14, 15, 17, 18, 20, 25, 29, 30, 31, 32, 40, 50, 75, 100 or 111 contiguous amino acids of SEQ ID NO:2 and can encode the entire amino acid sequence shown in SEQ ID NO:2. *A CSP56* polynucleotide can comprise a contiguous sequence of at least 10, 11, 12, 15, 20, 24, 25, 30, 32, 33, 35, 36, 40, 42, 45, 48, 50, 51, 54, 60, 63, 69, 70, 74, 75, 80, 84, 87, 90, 93, 96, 99, 100, 105, 114, 120, 125, 150, 225, 300, 333, or 336 nucleotides selected from SEQ ID NO:1 or can comprise SEQ ID NO:1. Preferred polynucleotides encode at least amino acids 1-30, 8-20, 7-21, 6-22, 106-115, 105-116, 104-117, 100-120, 297-306, 296-307, 295-308, 290-320, 461-489, 460-490, 459-491, and 407-518 of SEQ ID NO:2.

The complement of the nucleotide sequence shown in SEQ ID NO:1 is a contiguous nucleotide sequence which forms Watson-Crick base pairs with a contiguous nucleotide sequence shown in SEQ ID NO:1. The complement of SEQ ID NO:1 is a polynucleotide of the invention and can be used to provide *CSP56* antisense oligonucleotides and probes. Antisense oligonucleotides and probes of the invention can consist of at least 11, 12, 15, 20, 25, 30, 50, or 100 contiguous nucleotides which are complementary to the coding sequence shown in SEQ ID NO:1. A complement of the entire coding sequence can also be used. Double-stranded polynucleotides which comprise all or a portion of the nucleotide sequence shown in SEQ ID NO:1, as well as polynucleotides which encode CSP56-specific antibodies or ribozymes, are also polynucleotides of the invention.

Degenerate nucleotide sequences encoding amino acid sequences of CSP56

protein and or variants, as well as homologous nucleotide sequences which are at least 65%, 75%, 85%, 90%, 95%, 98%, or 99% identical to the nucleotide sequence shown in SEQ ID NO:1, are also *CSP56* polynucleotides. Percent sequence identity between the nucleotide sequence of SEQ ID NO:1 and a homologous or degenerate *CSP56* nucleotide sequence can be determined using computer programs which employ the Smith-Waterman algorithm, for example as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with the following parameters: a gap open penalty of 12 and a gap extension penalty of 1.

Typically, homologous *CSP56* sequences can be confirmed by hybridization under stringent conditions, as is known in the art. For example, using the following wash conditions--2X SSC, 0.1% SDS, room temperature twice, 30 minutes each; then 2X SSC, 0.1% SDS, 50 °C once for 30 minutes; then 2X SSC, room temperature twice, 10 minutes each--homologous sequences can be identified that contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15%, 2-10%, or 1-5% basepair mismatches. Degrees of homology of *CSP56* polynucleotides can be selected by varying the stringency of the wash conditions for identification of clones from gene libraries (or other sources of genetic material), as is well known in the art and described, for example, in manuals such as Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed. (1989). Species-specific homologs of *CSP56* polynucleotides of the invention can be identified by making suitable probes or primers and screening cDNA expression libraries from other species, such as mice, monkeys, yeast, or bacteria.

Nucleotide sequences which hybridize to the coding sequence shown in SEQ ID NO:1 or its complement following stringent hybridization and/or wash conditions are also *CSP56* subgenomic polynucleotides of the invention. Stringent wash conditions are well known and understood in the art and are disclosed, for example, in Sambrook *et al.*, 1989, at pages 9.50-9.51.

Typically, for stringent hybridization conditions a combination of temperature and salt concentration should be chosen that is approximately 12-20 °C below the

calculated T_m of the hybrid under study. It is well known that the T_m of a double-stranded DNA decreases by 1-1.5 °C with every 1% decrease in homology (Bonner *et al.*, *J. Mol. Biol. 81*, 123 (1973). The T_m of a hybrid between the *CSP56* sequence shown in SEQ ID NO:1 and a polynucleotide sequence which is 65%, 75%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:1 can be calculated, for example, using the equation of Bolton and McCarthy, *Proc. Natl. Acad. Sci. U.S.A. 48*, 1390 (1962):

$T_m = 81.5\ °C - 16.6(\log_{10}[Na^+]) + 0.41(\%G + C) - 0.63(\%formamide) - 600/l)$,

where l = the length of the hybrid in basepairs. Stringent wash conditions include, for example, 4X SSC at 65 °C, or 50% formamide, 4X SSC at 42 °C, or 0.5X SSC, 0.1% SDS at 65 °C. Highly stringent wash conditions include, for example, 0.2X SSC at 65 °C.

CSP56 polynucleotides can be purified free from other nucleotide sequences using standard nucleic acid purification techniques. For example, restriction enzymes and probes can be used to isolate polynucleotides which comprise nucleotide sequences encoding CSP56 protein. Alternatively, PCR can be used to synthesize and amplify such polynucleotides. At least 90% of a preparation of isolated and purified polynucleotides comprises *CSP56* encoding polynucleotides or their complement.

Complementary DNA (cDNA) molecules which encode CSP56 proteins are also *CSP56* subgenomic polynucleotides of the invention. *CSP56* cDNA molecules can be made with standard molecular biology techniques, using *CSP56* mRNA as a template. *CSP56* cDNA molecules can thereafter be replicated using molecular biology techniques known in the art and disclosed in manuals such as Sambrook *et al.*, 1989. An amplification technique, such as the polymerase chain reaction (PCR), can be used to obtain additional copies of subgenomic polynucleotides of the invention, using either human genomic DNA or cDNA as a template.

Alternatively, synthetic chemistry techniques can be used to synthesize *CSP56* subgenomic polynucleotide molecules of the invention. The degeneracy of the genetic code allows alternate nucleotide sequences to be synthesized which will encode a CSP56 protein having the amino acid sequence shown in SEQ ID NO:2 or a biologically active variant of that sequence. All such nucleotide sequences are within the scope of the

present invention.

The invention also provides polynucleotide probes which can be used to detect *CSP56* sequences, for example, in hybridization protocols such as Northern or Southern blotting or *in situ* hybridizations. Polynucleotide probes of the invention comprise at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, or 40 or more contiguous nucleotides selected from SEQ ID NO:1. Polynucleotide probes of the invention can comprise a detectable label, such as a radioisotopic, fluorescent, enzymatic, or chemiluminescent label.

Isolated *CSP56* polynucleotides can be used, for example, as primers to obtain additional copies of the polynucleotides or as probes for detecting *CSP56* mRNA. *CSP56* polynucleotides can also be used to express *CSP56* mRNA, protein, polypeptides, biologically active variants, single-chain antibodies, ribozymes, or fusion proteins.

Any of the *CSP56* polynucleotides described above can be present in a construct, such as a DNA or RNA construct. The construct can be a vector and can be used to transfer *a CSP56* polynucleotide into a cell, for example, for propagation of the polynucleotide. Constructs can be linear or circular molecules. They can be on autonomously replicating molecules or on molecules without replication sequences, and they can be regulated by their own or by other regulatory sequences, as is known in the art.

A construct can also be an expression construct. *A CSP56* expression construct comprises a promoter which is functional in a selected host cell. For example, the skilled artisan can readily select an appropriate promoter from the large number of cell type-specific promoters known and used in the art. The expression construct can also contain a transcription terminator which is functional in the host cell. The expression construct comprises a polynucleotide segment which encodes, for example, all or a portion of a CSP56 protein, polypeptide, biologically active variant, antibody, ribozyme, or fusion protein. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter. The expression construct can be linear or circular and can contain sequences, if desired, for autonomous replication.

Host cells which comprise any of the constructs described above can be

constructed using standard molecular biology techniques. Host cells of the invention can be used for a variety of purposes, such as propagation or expression of *CSP56* polynucleotides of the invention or for various assays. Host cells comprising constructs of the invention can be prokaryotic or eukaryotic. For example, bacterial, yeast, insect, mammalian, or human cells can be used to construct recombinant host cells.

Polynucleotides or constructs can be introduced into host cells using any technique known in the art. These techniques include transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and calcium phosphate-mediated transfection.

The *CSP56* gene is over-expressed in tumors, particularly in tumors with high metastatic potential compared with expression levels of *CSP56* in normal tissue or in tumors with low metastatic potential. The expression pattern of CSP56 suggests that this protease might be involved in a later step in the pathogenesis of cancer, particularly cancer of the breast and colon. The expression pattern of CSP56 suggests that this protease might be involved in a later step in the pathogenesis of cancer, particularly cancer of the breast and colon. Expression products of the *CSP56* gene can therefore be measured in a tumor sample in order to diagnose or prognose tumors with a high probability of metastasizing. CSP56 protein can also be measured in samples of a tumor over time, for example, to determine if the metastatic potential of a tumor has changed in response to a particular treatment. The *CSP56* gene is over-expressed if it is expressed at least 0.25-, 0.5-, 1-, 1.5-, 2-, or 3-fold higher in one tumor sample compared with expression levels in a normal tissue or in another tumor sample with a low metastatic potential.

Either CSP56 protein or mRNA can be measured in a body sample, particularly a breast or a colon sample, and compared with normal tissue from the same or a different body, preferably a human. Protein levels can be measured, for example, using CSP56-specific antibodies to detect CSP56 protein in tissue sections or homogenates of the tumor sample. Antibodies which specifically bind to CSP56 protein can be prepared as

described above and can comprise a detectable label, such as a radioisotope or a biotinylated, fluorescent, or chemiluminescent label. Any immunoassay known in the art can be used to detect CSP56 protein using CSP56-specific antibodies of the invention.

CSP56 mRNA can be measured, *inter alia*, using nucleotide probes which specifically hybridize to *CSP56* mRNA in assays such as Northern or dot blots or in an *in situ* hybridization protocol. Nucleotide probes can also comprise detectable labels such as those disclosed above. Appropriate nucleotide probes can be selected from the complement of the nucleotide sequence shown in SEQ ID NO:1.

If an expression product of the *CSP56* gene is detected in the body sample, the body sample is identified as neoplastic. If the expression product is not detected, the body sample is identified as normal. Furthermore, neoplastic potential of a tumor can be assessed by determining relative expression levels of *CSP56* mRNA or protein. High expression levels of *CSP56* mRNA or protein in a tumor sample indicate that the tumor has high metastatic potential. If very low levels of the expression product are detected, however, the tumor is identified as having a low potential for metastasizing.

Metastasis of a tumor, particularly a breast or a colon tumor, can be suppressed by contacting the tumor with a reagent which specifically binds to an expression product of *CSP56*. In one embodiment of the invention, expression of *CSP56* is decreased using a ribozyme, an RNA molecule with catalytic activity. *See, e.g.*, Cech, 1987, *Science 236*: 1532-1539; Cech, 1990, *Ann. Rev. Biochem. 59*:543-568; Cech, 1992, *Curr. Opin. Struct. Biol. 2*: 605-609; Couture and Stinchcomb, 1996, *Trends Genet. 12*: 510-515. Ribozymes can be used to inhibit gene function by cleaving an RNA sequence, as is known in the art (*e.g.*, Haseloff *et al.*, U.S. 5,641,673).

The coding sequence shown in SEQ ID NO:1 can be used to generate a ribozyme which will specifically bind to *CSP56* mRNA. Methods of designing and constructing ribozymes which can cleave other RNA molecules in trans in a highly sequence specific manner have been developed and described in the art (see Haseloff *et al.*, *Nature 334*:585-591, 1988). For example, the cleavage activity of ribozymes can be targeted to specific RNAs by engineering a discrete "hybridization" region into the ribozyme. The hybridization region contains a sequence complementary to the target RNA and thus

specifically hybridizes with the target (see, for example, Gerlach *et al.*, EP 321,201). Longer complementary sequences can be used to increase the affinity of the hybridization sequence for the target. The hybridizing and cleavage regions of the ribozyme can be integrally related; thus, upon hybridizing to the target RNA through the complementary regions, the catalytic region of the ribozyme can cleave the target.

CSP56 ribozymes can be introduced into cells as part of a DNA construct, as is known in the art. The DNA construct can also include transcriptional regulatory elements, such as a promoter element, an enhancer or UAS element, and a transcriptional terminator signal, for controlling transcription of the ribozyme in the cells.

Mechanical methods, such as microinjection, liposome-mediated transfection, electroporation, or calcium phosphate precipitation, can be used to introduce the *CSP56* ribozyme-containing DNA construct into cells in order to decrease *CSP56* expression. Alternatively, if it is desired that the cells stably retain the DNA construct, it can be supplied on a plasmid and maintained as a separate element or integrated into the genome of the cells, as is known in the art.

Expression of *CSP56* can also be altered using an antisense oligonucleotide. The sequence of the antisense oligonucleotide is complementary to at least a portion of the coding sequence shown in SEQ ID NO:1. Preferably, the antisense oligonucleotide is at least six nucleotides in length, but can be at least 8, 11, 12, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides long. Longer sequences, such as the complement of the nucleotide sequence shown in SEQ ID NO:1, can also be used. Antisense oligonucleotides can be provided in *a CSP56* construct of the invention and introduced into tumor cells, using transfection techniques known in the art.

CSP56 antisense oligonucleotides can be composed of deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides can be synthesized manually or by an automated synthesizer, by covalently linking the 5' end of one nucleotide with the 3' end of another nucleotide with non-phosphodiester internucleotide linkages such alkylphosphonates, phosphorothioates, phosphorodithioates, alkylphosphonothioates, alkylphosphonates, phosphoramidates, phosphate esters, carbamates, acetamidate, carboxymethyl esters, carbonates, and phosphate triesters. *See* Brown, 1994, *Meth. Mol.*

Biol. 20:1-8; Sonveaux, 1994, *Meth. Mol. Biol. 26*:1-72; Uhlmann *et al.*, 1990, *Chem. Rev. 90*:543-583.

Although precise complementarity is not required for successful duplex formation between *a CSP56* antisense oligonucleotide and the complementary coding sequence of *CSP56*, antisense oligonucleotides with no more than one mismatch are preferred. One skilled in the art can easily use the calculated melting point of *a CSP56* antisense-sense pair to determine the degree of mismatching which will be tolerated between a particular antisense oligonucleotide and a particular coding sequence of *CSP56*.

CSP56 antisense oligonucleotides can be modified without affecting their ability to hybridize to *a CSP56* coding sequence. These modifications can be internal or at one or both ends of the antisense oligonucleotide. For example, internucleoside phosphate linkages can be modified by adding cholesteryl or diamine moieties with varying numbers of carbon residues between the amino groups and terminal ribose. Modified bases and/or sugars, such as arabinose instead of ribose, or a 3', 5'-substituted oligonucleotide in which the 3' hydroxyl group or the 5' phosphate group are substituted, can also be employed in a modified antisense oligonucleotide. These modified oligonucleotides can be prepared by methods well known in the art. Agrawal *et al.*, *Trends Biotechnol. 10*:152-158, 1992; Uhlmann *et al.*, *Chem. Rev. 90*:543-584, 1990; Uhlmann *et al.*, *Tetrahedron. Lett. 215*:3539-3542, 1987.

Antibodies of the invention which specifically bind to CSP56 protein can also be used to alter expression of *CSP56*. Specific antibodies bind to CSP56 protein and prevent the protein from functioning in the cell. For example, polynucleotides encoding single-chain antibodies of the invention can be introduced into cells, using standard transfection techniques. Alternatively, therapeutic antibodies of the invention can be targeted to a particular cell type, for example, by binding an antibody to a coupling molecule which is specific for both the antibody and the target, as disclosed in WO 95/08577. The coupling molecule can comprise immunoglobulin binding domains.

Receptor-mediated targeted delivery of therapeutic compositions containing antibodies of the invention can also be used to deliver the antibodies to specific tissues. For example, many tumors, including breast, lung, and ovarian carcinomas, overexpress

antigens specific to malignant cells, such as glycoprotein $p185^{HER2}$. Antibodies which specifically bind to these antigens can be bound to liposomes which contain a CSP56 antibody of the invention. When injected into the bloodstream of a patient, the anti-$p185^{HER2}$ antibody directs the liposomes to the target cancer cells, where the liposomes are endocytosed and thus deliver their contents to the neoplastic cell (see Kirpotin *et al.*, *Biochem. 36*: 66, 1997).

In a preferred embodiment, a $p185^{HER2}$ antibody targeted delivery system is used to deliver an antibody which specifically binds to a CSP56 protein in a cancer cell. Liposomes can be loaded with the antibody as is known in the art (see Papahadjopoulos *et al.*, *Proc. Natl. Acad. Sci. 88*: 11640, 1991; Gabizon, *Cancer Res. 52*: 891, 1992; Lasic and Martin, *Stealth Liposomes*, 1995; Lasic and Papahadjopoulos, *Science 267*: 1275, 1995; and Park *et al.*, *Proc. Natl. Acad. Sci. 92*: 1327, 1995).

Antibodies which specifically bind to CSP56 protein, *CSP56* antisense oligonucleotides, or *CSP56* polynucleotides which encode single-chain antibodies or ribozymes can be provided to a tumor in a therapeutic composition. Therapeutic compositions of the invention also comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are well known to those in the art. Such carriers include, but are not limited to, large, slowly metabolized macromolecules, such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Pharmaceutically acceptable salts can also be used in the composition, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as the salts of organic acids such as acetates, proprionates, malonates, or benzoates.

Therapeutic compositions can also contain liquids, such as water, saline, glycerol, and ethanol, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes, such as those described in U.S. 5,422,120, WO 95/13796, WO 91/14445, or EP 524,968 B1, can also be used as a carrier for the therapeutic composition.

Typically, a therapeutic CSP56 composition is prepared as an injectable, either as a liquid solution or suspension; however, solid forms suitable for solution in, or

suspension in, liquid vehicles prior to injection can also be prepared. A CSP56 composition can also be formulated into an enteric coated tablet or gel capsule according to known methods in the art, such as those described in U.S. 4,853,230, EP 225,189, AU 9,224,296, and AU 9,230,801.

Administration of therapeutic compositions of the invention can include local or systemic administration, including injection, oral administration, particle gun, or catheterized administration, and topical administration. In addition, various methods can be used to administer a therapeutic CSP56 composition directly to a specific site in the body. For example, a small tumor or a metastatic lesion can be located and a CSP56 composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor can be identified, and a therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor.

A tumor which has a necrotic center can be aspirated, and the composition can be injected directly into the now empty center of the tumor. A therapeutic CSP56 composition can be directly administered to the surface of a tumor, for example, by topical application of the composition. X-ray imaging can be used to assist in certain of the above delivery methods. Combination therapeutic agents, including *a CSP56* polynucleotide and other therapeutic agents, can be administered simultaneously or sequentially.

Receptor-mediated targeted delivery can also be used to deliver therapeutic CSP56 compositions to specific tissues. Receptor-mediated delivery techniques are described in, for example, Findeis *et al.*, *Trends in Biotechnol. 11*, 202-05 (1993); Chiou *et al.*, GENE THERAPEUTICS: METHODS AND APPLICATIONS OF DIRECT GENE TRANSFER (J.A. Wolff, ed.) (1994); Wu & Wu (1988), *J. Biol. Chem. 263*, 621-24; Wu *et al.*, *J. Biol. Chem. 269*, 542-46 (1994); Zenke *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 87*, 3655-59 (1990); Wu *et al.*, *J. Biol. Chem. 266*, 338-42 (1991).

Both the dose of a CSP56 composition and the means of its administration can be determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors.

Preferably, a therapeutic composition of the invention decreases the level of CSP56 protein in the tumor by at least 50%, 60%, 70%, or 80%. Most preferably, the level of CSP56 protein in the tumor is decreased by at least 90%, 95%, 99%, or 100%. The effectiveness of the therapeutic composition can be assessed using methods well known in the art, such as hybridization of nucleotide probes to CSP56 mRNA, quantitative RT-PCR, or detection of CSP56 protein, using specific antibodies of the invention.

If the composition contains CSP56 antibodies, effective dosages of the composition are in the range of about 5 µg to about 50 µg/kg of patient body weight, about 50 µg to about 5 mg/kg, about 100 µg to about 500 µg/kg of patient body weight, and about 200 to about 250 µg/kg. Therapeutic compositions containing *CSP56* polynucleotides can be administered in a range of about 100 ng to about 200 mg of DNA for local administration. Concentration ranges of about 500 ng to about 50 mg, about 1 µg to about 2 mg, about 5 µg to about 500 µg, and about 20 µg to about 100 µg of DNA can also be used.

Factors such as method of action and efficacy of transformation and expression are considerations that will affect the dosage required for ultimate efficacy of the *CSP56* polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of *CSP56* polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

Expression of an endogenous *CSP56* gene in a cell can also be altered by introducing in-frame with the endogenous *CSP56* gene a DNA construct comprising *a* *CSP56* targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site by homologous recombination, such that a homologously recombinant cell comprising the DNA construct is formed. The new transcription unit can be used to turn the *CSP56* gene on or off as desired. This method of affecting endogenous gene expression is taught in U.S. Patent No. 5,641,670, which is incorporated herein by reference.

The targeting sequence is a segment of at least 10, 12, 15, 20, or 50 contiguous nucleotides selected from the nucleotide sequence shown in SEQ ID NO:1 or the complement thereof. The transcription unit is located upstream of a coding sequence of the endogenous *CSP56* gene. The exogenous regulatory sequence directs transcription of the coding sequence of the *CSP56* gene.

The invention also provides methods of screening test compounds for therapeutic effects. For example, synthesis of CSP56 protein can be measured to screen test compounds for the ability to suppress the metastatic potential of a tumor. The test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity. The compounds can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants. Test compounds can be produced recombinantly or synthesized by chemical methods known in the art.

A cell can be contacted with a test compound. Any cell which is capable of synthesizing CSP56 protein and which can be maintained *in vitro*, such as MDA-MB-435 cells, is suitable for use in this method. Synthesis of CSP56 protein can be measured by any means for measuring protein synthesis known in the art, such as incorporation of labeled amino acids into CSP56 protein followed by detection of labeled CSP56 protein in a polyacrylamide gel or a cell lysate. The amount of CSP56 protein can be detected, for example, using CSP56 protein-specific antibodies in Western blots. The amount of CSP56 protein synthesized in the presence or absence of a test compound can also be determined by comparing the amount of CSP56 protein synthesized with the amount of the CSP56 protein present in a standard curve.

Typically, a cell is contacted with a range of concentrations of the test compound, such as 1.0 nM, 5.0 nM, 10 nM, 50 nM, 100 nM, 500 nM, 1 mM, 10 mM, 50 mM, and 100 mM. Preferably, the test compound increases or decreases the level of CSP56 protein by at least 60%, 75%, or 80%. More preferably, an increase or decrease of at least 85%, 90%, 95%, or 98% is achieved. A test compound which increases the amount of CSP56 protein synthesized in the cell is identified as an agent which will increase the metastatic potential of a tumor. A test compound which decreases the amount of CSP56

protein synthesized in the cell is identified as potential therapeutic agent for decreasing metastatic potential of a tumor.

All of the references cited in this specification are expressly incorporated by reference in this disclosure. The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples which are provided herein for purposes of illustration only and are not intended to limit the scope of the invention.

EXPERIMENTAL PROCEDURES

The following materials and methods were used in the examples below.

Cell lines. Cell lines MCF-7, BR-3, BT-20, ZR-75-1, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-435, MDA-MB-453, MDA-MB-468, Alab, and Hs578Bst were obtained from American Type Culture Collection. All cell lines were grown according to their specifications.

Differential Display. Differential display was performed using the Hieroglyph mRNA profile kit according to the manufacturer's directions (Genomyx Corp., Foster City, CA). A total of 200 primer pairs were used to profile gene expression. Following amplification of randomly primed mRNAs by reverse-transcription-polymerase chain reaction (RT-PCR), the cDNA products were separated on 6% sequencing-type gels using a genomyxLR sequencer (Genomyx Corp.). The dried gels were exposed to Kodak XAR-2 film (Kodak, Rochester, NY) for various times.

Differentially-expressed cDNA fragments were excised and reamplified according to the manufacturer's directions (Genomyx Corp.). Because a gel slice excised from the gel contains 1 to 3 cDNA fragments of the same size (Martin *et al.*, *BioTechniques 24*, 1018-26, 1998; Giese *et al.*, Differential Display, Academic Press, 1998), reamplified products were separated by single strand confirmation polymorphism gels as described in (Mathieu-Dande *et al.*, *Nucl. Acids Res. 24*, 1504-07, 1996) and directly sequenced using M13 universal and T7 primers.

Construction and screening of human bone marrow stromal cell cDNA library. RNA was isolated from human bone marrow stromal cells (Poietic Technologies, Inc.,

Germantown, MD) using a guanidinium thiocyanate/phenol chloroform extraction protocol (Chirgwin *et al.*, *Biochem. 18*, 5294-99, 1979). Poly(A)$^+$ RNA was isolated using oligo-dT spin columns (Stratagene, La Jolla, CA). First and second strand synthesis was carried out according to the manufacturer's instructions (Pharmacia, Piscataway, NJ). Double-stranded cDNA was ligated into pBK-CMV phagemid vector (Stratagene, La Jolla, CA). Approximately, 1 x 10^6 plaques were screened using a 1.2 kb CSP56 cDNA fragment. Plasmid DNA from positive clones was obtained according to the manufacturer's instructions. Correctness of the nucleotide sequence was determined by double-strand sequencing.

Northern blot analysis and RT-PCR. Northern blots containing poly(A)$^+$ RNA prepared from various human normal and tumor tissues were purchased from ClonTech (Palo Alto, CA) and Biochain Institute (San Leandro, CA). All other Northern blots were prepared using 20 to 30 µg total RNA isolated using a guanidinium thiocyanate/phenol chloroform extraction protocol (Chirgwin *et al.*, 1979) from different human breast cancer and normal cell lines. Northern blots were hybridized at 65 °C in Express-hyb (ClonTech).

RT-PCR was performed using the reverse transcriptase RNA PCR kit (Perkin-Elmer, Roche Molecular Systems, Inc., Branchburg, NJ) according to the manufacturer's instructions.

In situ hybridization. In situ hybridization was performed on human tissues, frozen immediately after surgical removal and cryosection at 10 µm, following the protocol of Pfaff *et al.*, *Cell 84*, 309-20, 1996. Digoxigenin-UTP-labeled riboprobes were generated using the CSP56-containing plasmid DNA as a template. For generation of the antisense probe, the DNA was linearized with *Eco*RI (approximately 1 kb transcript) or *Nco*I (full-length transcript) and transcribed with T3 polymerase. For the sense control, the DNA was linearized with *Xho*I (full-length transcript) and transcribed with T7 polymerase. Hybridized probes were detected with alkaline phosphatase-coupled anti-digoxigenen antibodies using BM Purple as the substrate (Boehringer Mannheim).

Tumor growth in the mammary fatpad of immunodeficient mice. Scid (severe

combined immunodeficient) mice (Jackson Laboratory) were anesthetized, and a small incision was made to expose the mammary fatpad. Approximately 4 x 10^6 cells were injected into the fatpad of each mouse. Tumor growth was monitored by weekly examination, and growth was determined by caliper measurements. After approximately 4 weeks, primary tumors were removed from anesthetized mice, and the skin incisions were closed with wound clips. Approximately 4 weeks later, mice were killed and inspected for the presence of lung metastases. Primary tumors and lung metastasis were analyzed histologically for the presence of human cells. A chunk of tumor tissue representing more than 80% cells of human origin was used to isolate total RNA. In the case of MDA-MD-435, large lung metastases representing more than 90% human cells were used. Total RNA was amplified by RT-PCR using specific primers for the CSP56 coding region. The reaction products were dot blotted onto nylon membranes and hybridized with a CSP56-specific probe.

EXAMPLE 1

This example demonstrates identification of a differentially-expressed gene in the aggressive-invasive human breast cancer cell line MDA-MB-435.

To identify genes associated with the metastatic phenotype, we compared the gene expression profiles in four human breast cancer cell lines using which display different malignant phenotypes, MDA-MB-453, MCF-7, MDA-MB-231, and MDA-MB-435, ranging from poorly-invasive to most aggressively-invasive (Engel *et al.*, *Cancer Res. 38*, 4327-39, 1978; Shafie and Liotta, *Cancer Lett. 11*, 81-87, 1990; Ozello and Sordat, *Eur. J. Cancer 16*, 553-59, 1980; Price *et al.*, *Cancer Res. 50*, 717-21, 1990). Cell lines were chosen as starting material based on the ability to obtain high amounts of pure RNA. In contrast, human breast cancer biopsies consist of a mixture of cancer and other cell types including macrophages and lymphocytes (Kelly *et al.*, *Br. J. Cancer 57*, 174-77, 1988; Whitford *et al.*, *Br. J. Cancer 62*, 971-75, 1990). The described human breast cancer cell lines have been extensively studied in mouse models allowing one to functionally characterize identified candidate genes in tumor progression.

To ensure that the cell lines retained their original malignant properties after

prolonged passage in culture, we examined their potential to grow in scid mice and to form metastasis following injection into the mammary fatpad. Three of the four cell lines formed primary tumors, consistent with previous reports (Engel *et al.*, 1978; Shafie and Liotta, 1990; Ozello and Sordat, 1980; Price *et al.*, 1990). No primary tumor formation was detected with MDA-MB-453. In addition, mice injected with MDA-MB-231 and MDA-MB-435 developed lung metastases, with the highest incidence being detected using MDA-MB-435.

Next, we performed a differential display analysis using total RNA isolated from the breast cancer cell lines and a total of 200 different primer pair combinations. Among several differentially expressed transcripts, a 1.2-kb cDNA fragment was specifically amplified from the MDA-MB-435 RNA sample using the primer pair combination, Ap8 [5'-ACGACTCACTATAGG $GC(T)_{12}AA$] (SEQ ID NO:3) and Arp1 (5'-ACAATTTCACACAGGACGACTCCAAG) (SEQ ID NO:4) (Figure 1A, lanes 5 and 6). Weak expression was also detected in MDA-MB-231 (Figure 1A, lanes 1 and 2), whereas no signal was detected in the RNA samples isolated from MCF-7 and MDA-MB-453 (Figure 1A, lanes 3, 4, 7, and 8).

To confirm the expression pattern, the DNA fragment was isolated from the gel, reamplified, radiolabeled, and used as a hybridization probe in a Northern blot analysis of human breast cancer cell lines with different malignant phenotypes and a non-tumorigenic breast cell line (Figure 1B). The radioactive probe hybridized with similar intensity to two transcripts of approximately 2.0-kb and 2.5-kb in size in the MDA-MB-435 RNA sample (lane 9). Weak expression of these transcripts was detected in the poorly invasive human breast cell lines (lanes 2 and 3) or in the non-tumorigenic line Hs578Bst (lane 1). No signal was detected in MDA-MB-453 and MCF-7. These data show a restricted expression pattern of this gene to highly or moderately metastatic human breast cancer cell lines.

EXAMPLE 2

This example demonstrates the nucleotide sequence of CSP56 cDNA.

Comparison of the nucleotide sequence of CSP56 cDNA to public databases showed no significant homologies. To obtain more nucleotide sequence information, we screened a human bone marrow stromal cell cDNA library. One of the positive clones extended the original clone to 1855 nucleotides in length (Figure 2A). This sequence was further extended at the 3'-end with several expressed sequenced tags to 2606 nucleotides in length (Figure 2B). The additional 750 nucleotides are most probably the result of alternative poly-A site selection.

Analysis of the nucleotide sequence revealed a single open reading frame of 518 amino acids, beginning with a start codon for translation at nucleotide position 101 and terminating with a stop codon at nucleotide position 1655. A consensus Kozak sequence (Kozak, *Cell 44*, 283-92, 1986) around the start codon and the analysis of the codon usage (Wisconsin package, UNIX) suggests that this cDNA clone contains the entire coding region.

Translation of the open reading frame predicts a protein with a molecular mass of 56 kD. On the basis of its specific expression in the highly metastatic human breast cancer cell lines, the cDNA-encoded protein was termed CSP56 for cancer-specific protein 56-kd.

EXAMPLE 3

This example demonstrates that CSP56 is a novel aspartyl-type protease.

Comparison of the CSP56 open reading frame with proteins in public databases shows some homology to members of the pepsin family of aspartyl proteases (Figure 3). A characteristic feature of this protease family is the presence of two active centers which evolved by gene duplication (Davies, *Ann. Rev. Biophys. Biochem. 19*, 189-215, 1990; Neil and Barrett, *Meth. Enz. 248*, 105-80, 1995). The amino acid residues comprising the catalytic domains (Asp-Thr/Ser-Gly) and the flanking residues display the highest conservation in this family and are conserved in CSP56 (Figures 2 and 3).

CSP56, however, shows structural features which are distinct from other aspartyl

proteases. Overall similarities of CSP56 to pepsinogen C and A, renin, and cathepsin D and E are only 55, 51, 54, 52, and 51%, respectively, neglecting the CSP56 C-terminal extension. The cysteine residues found following and preceding the catalytic domains in other members are absent in CSP56 (Figure 3). CSP56 also contains a carboxy-terminal extension of approximately 90 amino acid residues which shows no significant homology to known proteins.

CSP56 also contains a hydrophobic motif consisting of 29 amino acid residues in the C-terminal extension which may function as a membrane attachment domain. (Figures 2C and 3) CSP56 also contains a putative signal sequence.

CSP56 is therefore a novel aspartyl-type protease with a putative transmembrane domain (amino acids 8-20) and a stretch of approximately 45 amino acids representing a putative propeptide (amino acids 21 to 76).

EXAMPLE 4

This example demonstrates the expression pattern of CSP56 throughout human breast cancer development and in metastasis.

To further examine the expression pattern of CSP56, we performed a Northern blot analysis using additional human breast cancer and normal cell lines (Figure 4). Expression of CSP56 was detected in MDA-MB-435, MDA-MB-468, and BR-3 (lanes 1, 4, and 9), with the strongest signal in MDA-MB-435. Other cell lines showed weak expression. No signal was detected in the poorly-invasive human breast cancer cell lines MDA-MB-453 and MCF-7 and in a normal breast cell line Hs578Bst. Together, these data are consistent with the increased expression of CSP56 in highly malignant human breast cancer cell lines.

EXAMPLE 5

This example demonstrates the expression pattern of CSP56 in normal human tissues.

To determine the tissue distribution of CSP56, polyA$^+$ RNA from various human tissues was examined by Northern blot analysis (Figure 7). Two major transcripts were

detected that are similar in size to those detected in cancer cell lines and human tissues. Highest expression was detected in pancreas, prostate, and placenta. Weak or no signal was detected in brain and peripheral blood lymphocytes.

EXAMPLE 6

This example demonstrates identification of CSP56 transcripts in primary tumors and metastatic lung tissue isolated from immunodeficient mice injected with MDA-MB-435.

The scid mouse model was used to examine CSP56 expression in tumors. This model has been shown to be suitable for evaluating the function of genes implicated in the tumorigenicity and metastasis of human breast cancer cells (Steeg *et al.*, *Breast Cancer Res. Treat. 25*, 175-87, 1993; Price, *Breast Cancer Res. Treat. 39*, 93-102, 1996).

Different human breast cancer cell lines were injected into the mammary fatpad of immunodeficient mice. Primary tumors and, if applicable, lung metastases were isolated from mice, and total RNA was prepared for Northern blot analysis (Figure 4).

CSP56 transcripts were detected in primary tumor RNA derived from MDA-MB-435, MDA-MB-468 and Alab, but not from MCF-7 (Figure 4). CSP56 gene expression was also detected in lung metastasis of mice injected with MDA-MB-435 (lane 1). Failure to detect CSP56 transcripts in primary tumors of mice injected with ZR-75-1, MDA-MB-361, and MDA-MB-231 could be explained with the small amount of human cancer tissues in these tumors as judged by the weak human β-actin signal when compared to other primary tumor RNA samples.

Together these data exclude *in vitro* culture conditions as a cause for CSP56 up-regulation and establishes this gene as a novel tumor maker.

EXAMPLE 7

This example demonstrates detection of CSP56 gene expression detected in patient samples.

CSP56 expression was examined in RNA samples isolated from patient tumor biopsies. A Northern blot containing total RNA from breast tumor tissue and normal

breast tissue from the same patient was hybridized with a CSP56-specific probe (Fig. 5A). CSP56 transcripts were detected in the tumor sample whereas no signal was detected in the normal breast RNA (lanes 1 and 2). Similarly, expression of CSP56 transcripts were up-regulated in two other breast cancer RNA samples when compared to a normal breast RNA control (Fig. 5B). Increased expression of *CSP56* was also detected in human colon cancer tissue when compared to normal colon tissue of the same patient.

To identify the cell types that express CSP56 transcripts *in vivo*, we performed an *in situ* hybridization analysis on tissue samples obtained from one breast cancer patient (Figure 6A-6F). A weak CSP56 signal was detected in the cells of the ducts of normal breast tissue (Figure 6B). In the primary tumor, CSP56 was highly expressed in the tumor cells but not in the surrounding lymphocytes (Figure 6E). No signal was detected using the sense probe (Figures 6C and 6F).

We also analyzed tissue samples obtained from two colon cancer patients (Figures 6G-6M) for CSP56 expression. No signal was detected in normal colon tissue (Figure 6H), whereas CSP56 transcripts were abundant in the tumor cells of both the primary colon tumor and the liver metastasis, and no expression was detected in the surrounding stroma (Figures 6K and 6M).

These data demonstrate that CSP56 is over-expressed in tumor cells of human cancer patients and may play a role in the development and progression of different types of tumors.

CLAIMS:

1. An isolated human CSP56 protein having an amino acid sequence which is at least 85% identical to SEQ ID NO:2, wherein percent identity is determined using a Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

2. The isolated human CSP56 protein of claim 1 which has the amino acid sequence shown in SEQ ID NO:2.

3. An isolated polypeptide comprising at least 8 contiguous amino acids as shown in SEQ ID NO:2.

4. The isolated polypeptide of claim 4 which is selected from the group consisting of at least amino acids 461-489 of SEQ ID NO:2, at least amino acids 106-115 of SEQ ID NO:2, at least amino acids 297-306 of SEQ ID NO:2, and at least amino acids 8-20 of SEQ ID NO:2.

5. A CSP56 fusion protein comprising a first protein segment and a second protein segment fused together by means of a peptide bond, wherein the first protein segment consists of at least 8 contiguous amino acids of a human CSP56 protein as shown in SEQ ID NO:2.

6. A preparation of antibodies which specifically bind to a human CSP56 protein having an amino acid sequence as shown in SEQ ID NO:2.

7. A cDNA molecule which encodes a human CSP56 protein having an amino acid sequence which is at least 85% identical to SEQ ID NO:2, wherein percent identity is determined using a Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

8. A cDNA molecule which encodes at least 8 contiguous amino acids of SEQ ID NO:2.

9. The cDNA molecule of claim 8 which encodes SEQ ID NO:2.

10. The cDNA molecule of claim 9 which comprises SEQ ID NO:1.

11. A cDNA molecule comprising at least 12 contiguous nucleotides of SEQ ID NO:1.

12. A cDNA molecule which is at least 85% identical to the nucleotide sequence shown in SEQ ID NO:1, wherein percent identity is determined using a Smith-Waterman homology search algorithm as implemented in a MPSRCH program using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 1.

13. An isolated and purified subgenomic polynucleotide comprising a nucleotide sequence which hybridizes to SEQ ID NO:1 after washing with 0.2X SSC at 65 °C, wherein the nucleotide sequence encodes a CSP56 protein having the amino acid sequence of SEQ ID NO:2.

14. A construct comprising:

a promoter; and

a polynucleotide segment encoding at least 8 contiguous amino acids of a human CSP56 protein as shown in SEQ ID NO:2, wherein the polynucleotide segment is located downstream from the promoter, wherein transcription of the polynucleotide segment initiates at the promoter.

15. A host cell comprising a construct which comprises:

a promoter and:

a polynucleotide segment encoding at least 8 contiguous amino acids of a human CSP56 protein having an amino acid sequence as shown in SEQ ID NO:2.

16. A recombinant host cell comprising a new transcription initiation unit, wherein the new transcription initiation unit comprises in 5' to 3' order:

(a) an exogenous regulatory sequence;

(b) an exogenous exon; and

(c) a splice donor site,

wherein the new transcription initiation unit is located upstream of a coding sequence of an *CSP56* gene as shown in SEQ ID NO:1, wherein the exogenous regulatory sequence controls transcription of the coding sequence of the *CSP56* gene.

17. A polynucleotide probe comprising at least 12 contiguous nucleotides of SEQ ID NO:1.

18. The polynucleotide probe of claim 17 which comprises a detectable label.

19. A method of diagnosing neoplasia, comprising the step of:

detecting in a body sample an expression product of the nucleotide sequence shown in SEQ ID NO:1, wherein detection of the expression product identifies the body sample as neoplastic.

20. The method of claim 19 wherein the body sample is selected from the group consisting of a breast sample and a colon sample.

21. The method of claim 19 wherein the expression product is protein.

22. The method of claim 21 wherein the protein is detected using an antibody which specifically binds to the protein.

23. The method of claim 19 wherein the expression product is mRNA.

24. The method of claim 23 wherein the mRNA is detected using a nucleotide probe which specifically hybridizes to the mRNA.

25. A method for determining metastatic potential of a tumor, comprising the step of:

measuring in a tumor sample an expression product of a gene having the coding sequence shown in SEQ ID NO:1, wherein a tumor sample which expresses the expression product is categorized as having metastatic potential.

26. The method of claim 25 wherein the expression product is protein.

27. The method of claim 26 wherein the protein is measured using an antibody which specifically binds to the protein.

28. The method of claim 25 wherein the expression product is mRNA.

29. The method of claim 28 wherein the mRNA is measured using a nucleotide probe which specifically hybridizes to the mRNA.

30. The method of claim 25 wherein the tumor is selected from the group consisting of a breast tumor and a colon tumor.

31. A method of screening test compounds for the ability to suppress the metastatic potential of a tumor, comprising the steps of:

contacting a cell with a test compound; and

measuring in the cell the synthesis of a protein having the amino acid sequence shown in SEQ ID NO:2, wherein a test compound which decreases the amount of the protein synthesized in the cell is identified as a potential agent for suppressing the

metastatic potential of the tumor.

32. A set of primers for amplifying at least a portion of a gene having the coding sequence shown in SEQ ID NO:1.

33. The set of claim 32 wherein the primers are the nucleotide sequences shown in SEQ ID NOs:3 and 4.

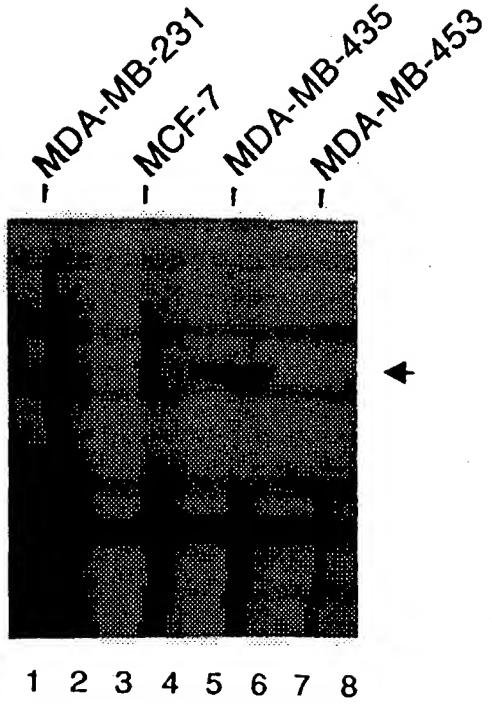

MDA-MB-231
MCF-7
MDA-MB-435
MDA-MB-453
1 2 3 4 5 6 7 8

FIG. 1A

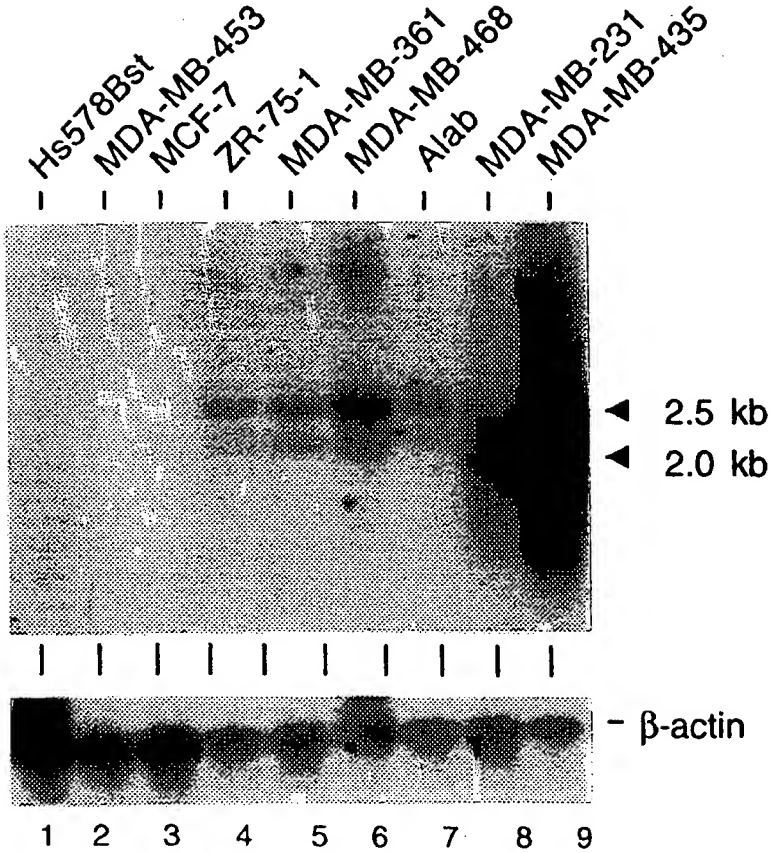

Hs578Bst
MDA-MB-453
MCF-7
ZR-75-1
MDA-MB-361
MDA-MB-468
Alab
MDA-MB-231
MDA-MB-435
2.5 kb
2.0 kb
β-actin
1 2 3 4 5 6 7 8 9

FIG. 1B

FIG. 2A-1

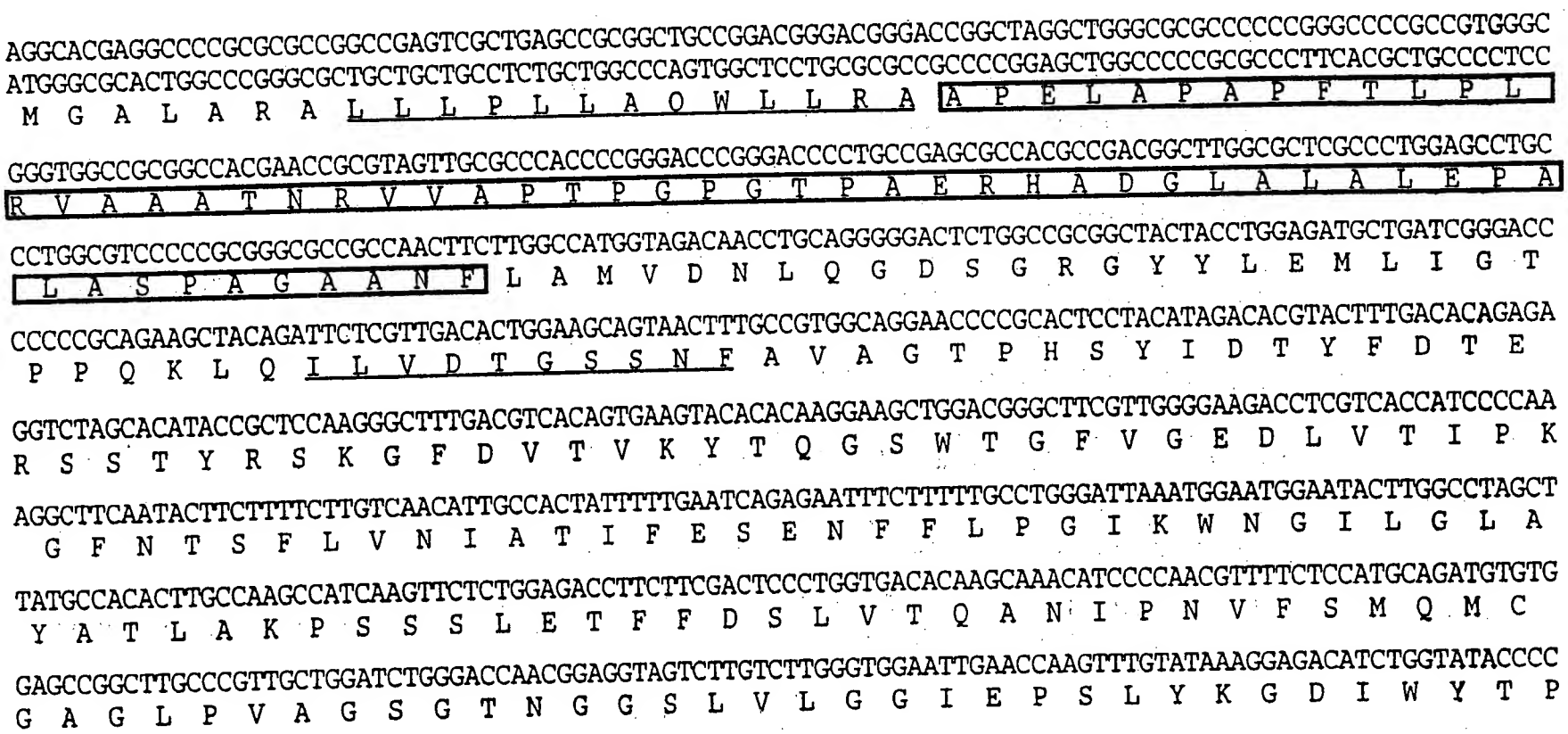

AGGCACGAGGCCCCGCGCGCCGGCCGAGTCGCTGAGCCGCGGCTGCCGGACGGGACGGGACCGGCTAGGCTGGGCGCGCCCCCCGGGCCCCGCCGTGGGC 100
ATGGGCGCACTGGCCCGGGCGCTGCTGCTGCCTCTGCTGGCCCAGTGGCTCCTGCGCGCCGCCCCGGAGCTGGCCCCCGCGCCCTTCACGCTGCCCCTCC 200
M G A L A R A L L L P L L A Q W L L R A A P E L A P A P F T L P L 33
GGGTGGCCGCGGCCACGAACCGCGTAGTTGCGCCCACCCCGGGACCCGGGACCCCTGCCGAGCGCCACGCCGACGGCTTGGCGCTCGCCCTGGAGCCTGC 300
R V A A A T N R V V A P T P G P G T P A E R H A D G L A L A L E P A 66
CCTGGCGTCCCCCGCGGGCGCCGCCAACTTCTTGGCCATGGTAGACAACCTGCAGGGGGACTCTGGCCGCGGCTACTACCTGGAGATGCTGATCGGGACC 400
L A S P A G A A N F L A M V D N L Q G D S G R G Y Y L E M L I G T 99
CCCCCGCAGAAGCTACAGATTCTCGTTGACACTGGAAGCAGTAACTTTGCCGTGGCAGGAACCCCGCACTCCTACATAGACACGTACTTTGACACAGAGA 500
P P Q K L Q I L V D T G S S N F A V A G T P H S Y I D T Y F D T E 132
GGTCTAGCACATACCGCTCCAAGGGCTTTGACGTCACAGTGAAGTACACACAAGGAAGCTGGACGGGCTTCGTTGGGGAAGACCTCGTCACCATCCCCAA 600
R S S T Y R S K G F D V T V K Y T Q G S W T G F V G E D L V T I P K 165
AGGCTTCAATACTTCTTTTCTTGTCAACATTGCCACTATTTTTGAATCAGAGAATTTCTTTTTGCCTGGGATTAAATGGAATGGAATACTTGGCCTAGCT 700
G F N T S F L V N I A T I F E S E N F F L P G I K W N G I L G L A 198
TATGCCACACTTGCCAAGCCATCAAGTTCTCTGGAGACCTTCTTCGACTCCCTGGTGACACAAGCAAACATCCCCAACGTTTTCTCCATGCAGATGTGTG 800
Y A T L A K P S S S L E T F F D S L V T Q A N I P N V F S M Q M C 231
GAGCCGGCTTGCCCGTTGCTGGATCTGGGACCAACGGAGGTAGTCTTGTCTTGGGTGGAATTGAACCAAGTTTGTATAAAGGAGACATCTGGTATACCCC 900
G A G L P V A G S G T N G G S L V L G G I E P S L Y K G D I W Y T P 264

FIG. 2A-2

```
TATTAAGGAAGAGTGGTACTACCAGATAGAAATTCTGAAATTGGAAATTGGAGGCCAAAGCCTTAATCTGGACTGCAGAGAGTATAACGCAGACAAGGCC  1000
  I  K  E  E  W  Y  Y  Q  I  E  I  L  K  L  E  I  G  G  Q  S  L  N  L  D  C  R  E  Y  N  A  D  K  A  297

ATCGTGGACAGTGGCACCACGCTGCTGCGCCTGCCCCAGAAGGTGTTTGATGCGGTGGTGGAAGCTGTGGCCCGCGCATCTCTGATTCCAGAATTCTCTG  1100
  I  V  D  S  G  T  T  L  L  R  L  P  Q  K  V  F  D  A  V  V  E  A  V  A  R  A  S  L  I  P  E  F  S  330

ATGGTTTCTGGACTGGGTCCCAGCTGGCGTGCTGGACGAATTCGGAAACACCTTGGTCTTACTTCCCTAAAATCTCCATCTACCTGAGAGATGAGAACTC  1200
  D  G  F  W  T  G  S  Q  L  A  C  W  T  N  S  E  T  P  W  S  Y  F  P  K  I  S  I  Y  L  R  D  E  N  S  363

CAGCAGGTCATTCCGTATCACAATCCTGCCTCAGCTTTACATTCAGCCCATGATGGGGGCCGGCCTGAATTATGAATGTTACCGATTCGGCATTTCCCCA  1300
  S  R  S  F  R  I  T  I  L  P  Q  L  Y  I  Q  P  M  M  G  A  G  L  N  Y  E  C  Y  R  F  G  I  S  P  396

TCCACAAATGCGCTGGTGATCGGTGCCACGGTGATGGAGGGCTTCTACGTCATCTTCGACAGAGCCCAGAAGAGGGTGGGCTTCGCAGCGAGCCCCTGTG  1400
  S  T  N  A  L  V  I  G  A  T  V  M  E  G  F  Y  V  I  F  D  R  A  Q  K  R  V  G  F  A  A  S  P  C  429

CAGAAATTGCAGGTGCTGCAGTGTCTGAAATTTCCGGGCCTTTCTCAACAGAGGATGTAGCCAGCAACTGTGTCCCCGCTCAGTCTTTGAGCGAGCCCAT  1500
A  E  I  A  G  A  A  V  S  E  I  S  G  P  F  S  T  E  D  V  A  S  N  C  V  P  A  Q  S  L  S  E  P  I  462

TTTGTGGATTGTGTCCTATGCGCTCATGAGCGTCTGTGGAGCCATCCTCCTTGTCTTAATCGTCCTGCTGCTGCTGCCGTTCCGGTGTCAGCGTCGCCCC  1600
  L  W  I  V  S  Y  A  L  M  S  V  C  G  A  I  L  L  V  L  I  V  L  L  L  L  P  F  R  C  Q  R  R  P  495

CGTGACCCTGAGGTCGTCAATGATGAGTCCTCTCTGGTCAGACATCGCTGGAAATGAATAGCCAGGCCTGACCTCAAGCAACCATGAACTCAGCTATTAA  1700
  R  D  P  E  V  V  N  D  E  S  S  L  V  R  H  R  W  K  *                                              513

GAAAATCACATTTCCAGGGCAGCAGCCGGGATCGATGGTGGCGCTTTCTCCTGTGCCCACCCGTCTTCAATCTCTGTTCTGCTCCCAGATGCCTTCTAGA  1800
TTCACTGTCTTTTGATTCTTGATTTTCAAGCTTTCAAATCCTCCCTACTTCCAAG(A)n  1855
```

FIG. 2B

```
AAAAAAAAAACTTCATTCTAAACCAAAACAGAGTGGATTGGGCTGCAGGCTCTATGGGGTTCGTTATGCCAAAGTGTCTACATGTGCCACCAACATAAAA  1955
CAAAACCAAGCCTTGGCTCGTTCTCTTCTCTCTTCAATCTCTGGAAAAATAAGTACATATAGTTGATAACCCCTCTTAGCTTACAGGAAGCTTTTTGTAT  2055
TAATTGCCTTTGAGGTTATTTTCCGCCAGACCTCAACCTGGGTCAAAGTGGTACAGGAAGGCTTGCAGTATGATGGCAGGAGAATCAGCCTGGGGCCTGG  2155
GGATGTAACCAAGCTGTACCCTTGAGACCTGGAACCAGAGCCACAGGCCCCTTTTGTGGGTTTCTCTGTGCTCTGAATGGGAGCCAGAATTCACTAGGAG  2255
GTCATCAACCGATGGTCCTCACAAGCCTCTTCTGAAGATGGAAGGCCTTTTGCCCGTTGAGGTAGAGGGGAAGGAAATCTCCTCTTTTGTACCCAATACT  2355
TATGTTGTATTGTTGGTGCGAAAGTAAAAACACTACCTCTTTTGAGACTTTGCCCAGGGTCCTGTGCCTGGATGGGGGTGCAGGCAGCCTTGACCACGGC  2455
TGTTCCCCTCACCCAAAAGAATTATCATCCCAACAGCCAAGACCCAACAGGTGCTGAACTGTGCATCAACCAGGAAGAGTTCTATCCCCAAGCTGGCCAC  2555
TATCACATATGCTTACTCTTGCTTAAAATTAATAAATCATGTTTTGATGAG(A)n      2606
```

FIG. 2C

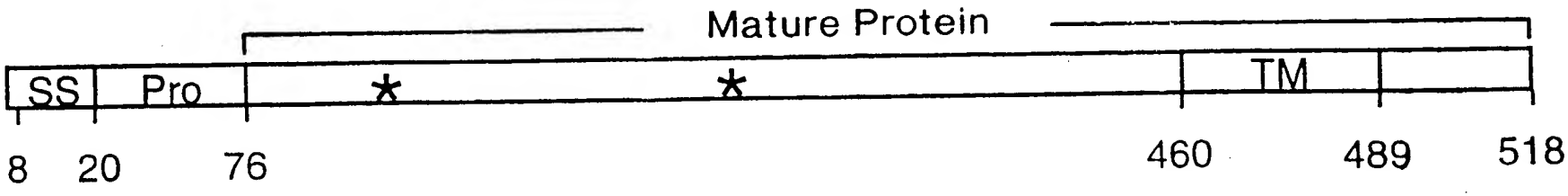


FIG. 3A

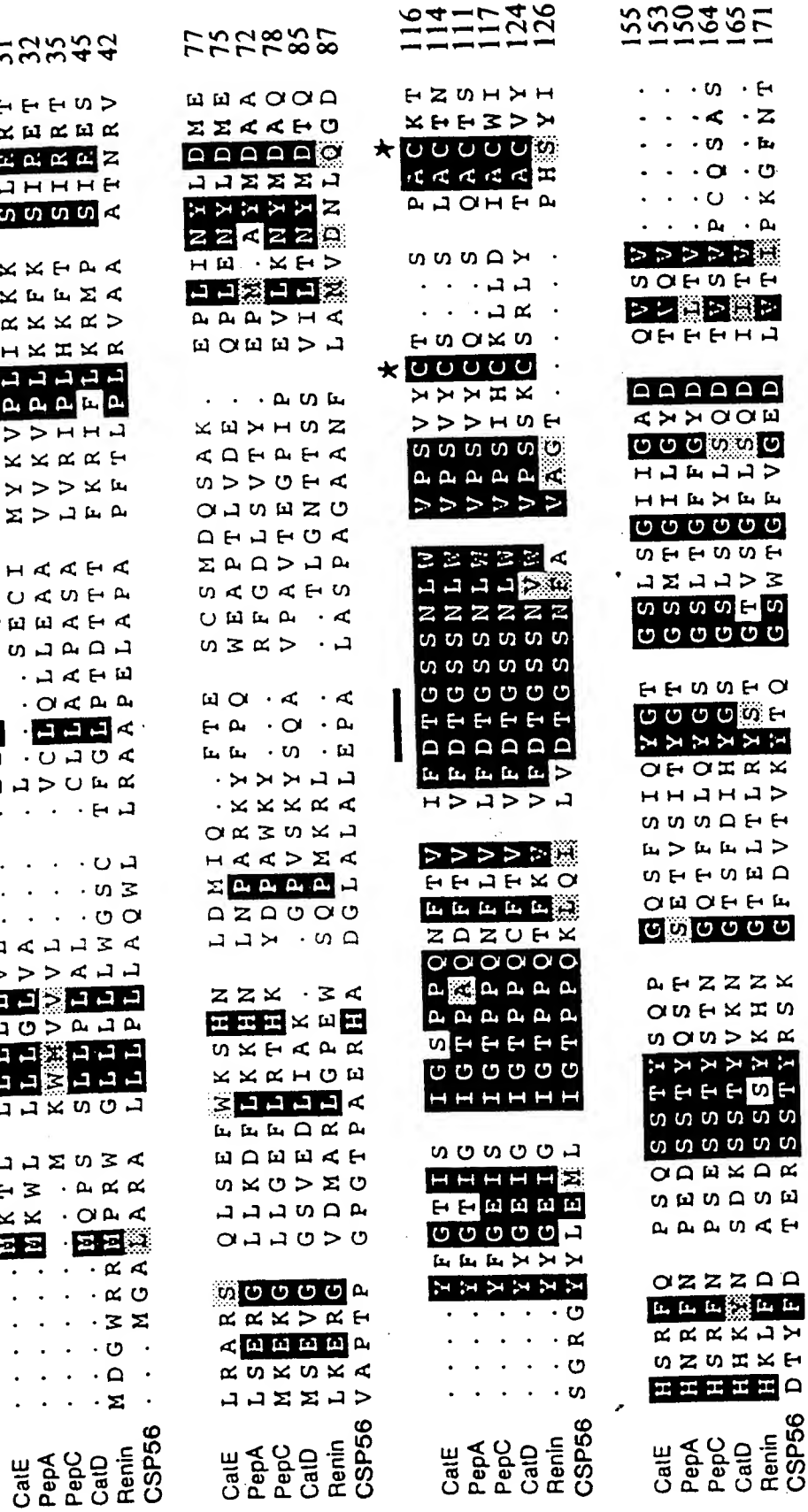

CatE
PepA
PepC
CatD
Renin
CSP56
35
31
32
35
45
42
77
75
72
78
85
87
116
114
111
117
124
126
155
153
150
164
165
171

FIG. 3B

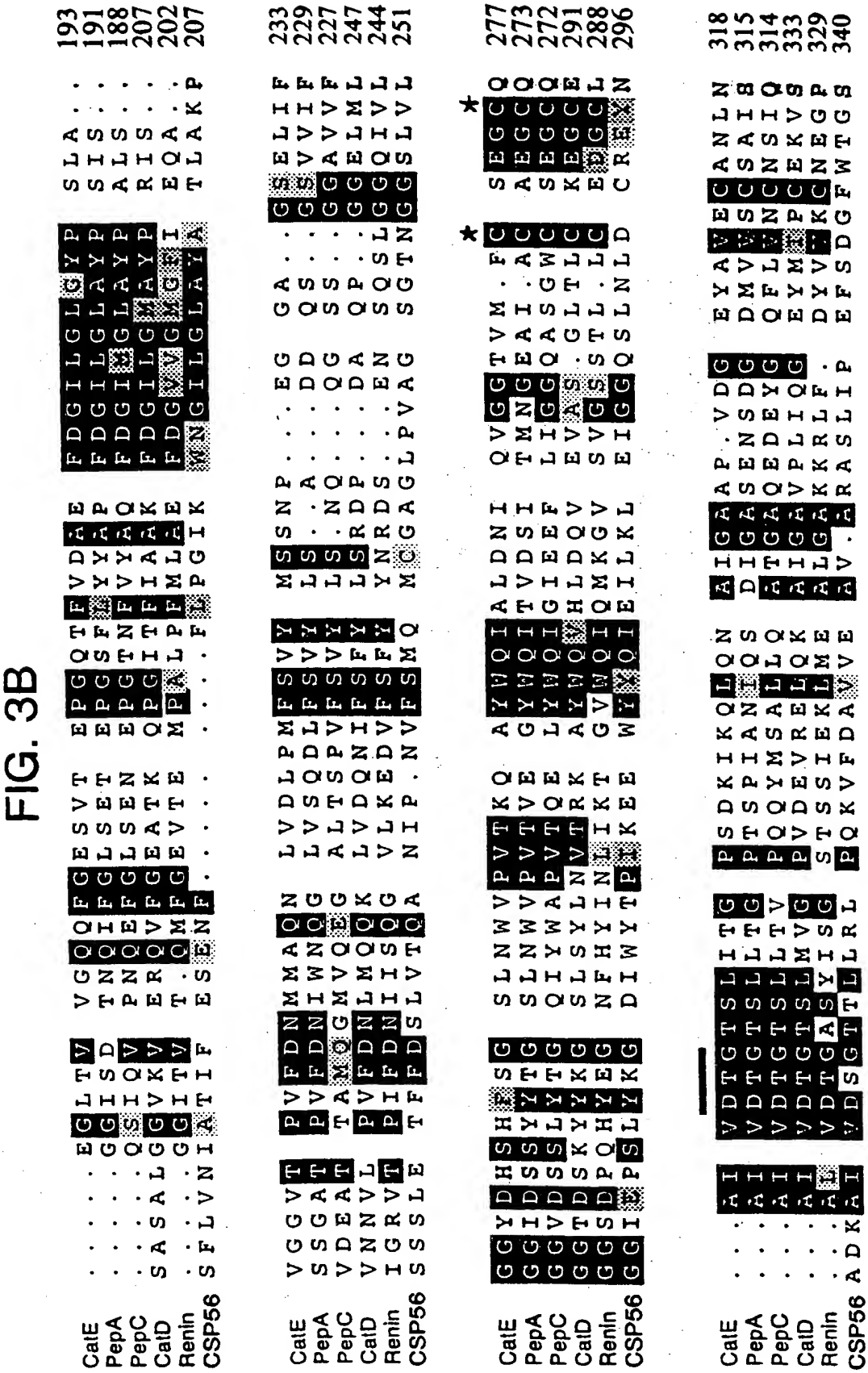
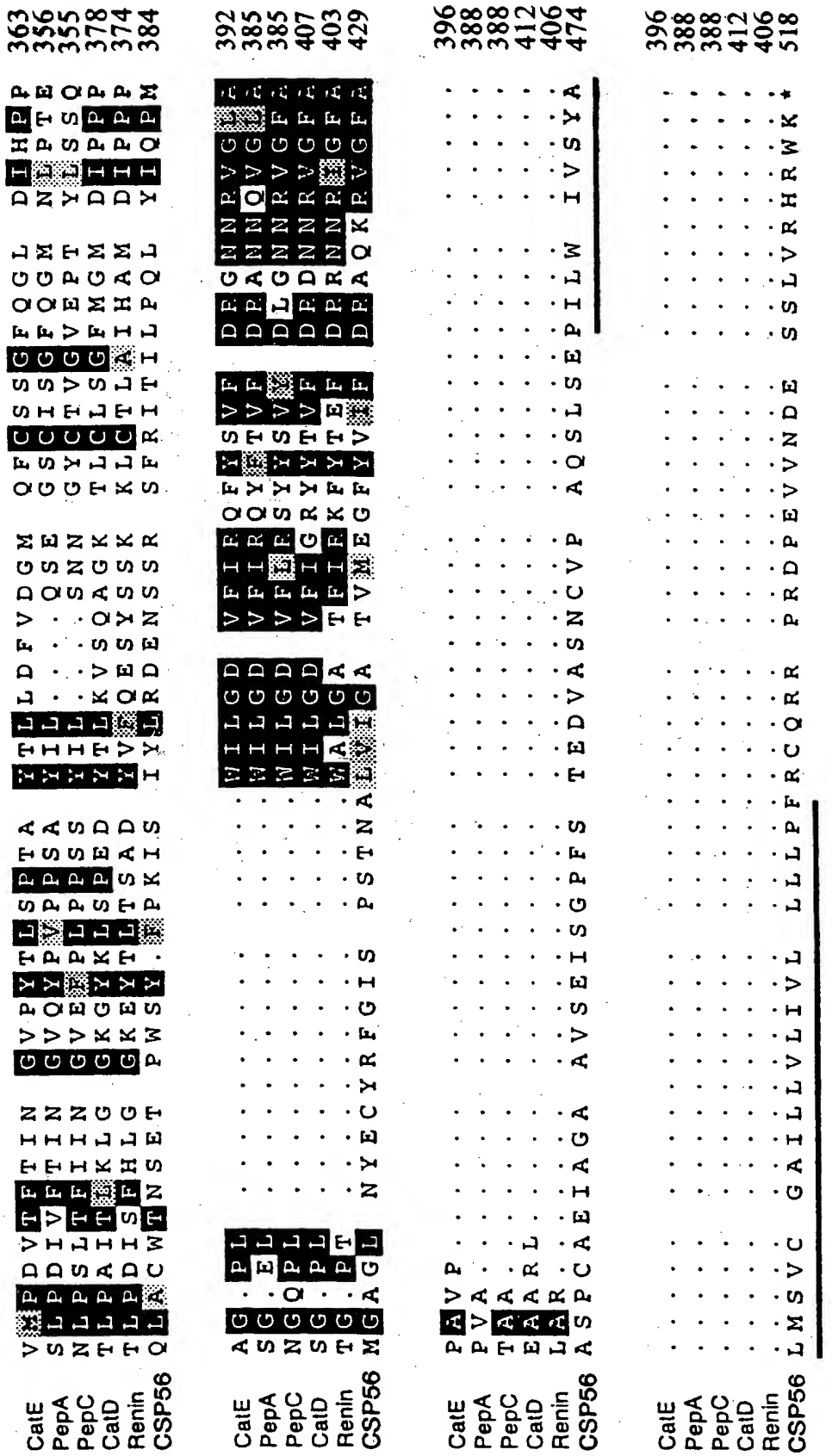

CatE
PepA
PepC
CatD
Renin
CSP56
193
191
188
207
202
207
233
229
227
247
244
251
277
273
272
291
288
296
318
315
314
333
329
340

FIG. 3C


CatE
PepA
PepC
CatD
Renin
CSP56
363
356
355
378
374
384
392
385
385
407
403
429
396
388
388
412
406
474
396
388
388
412
406
518

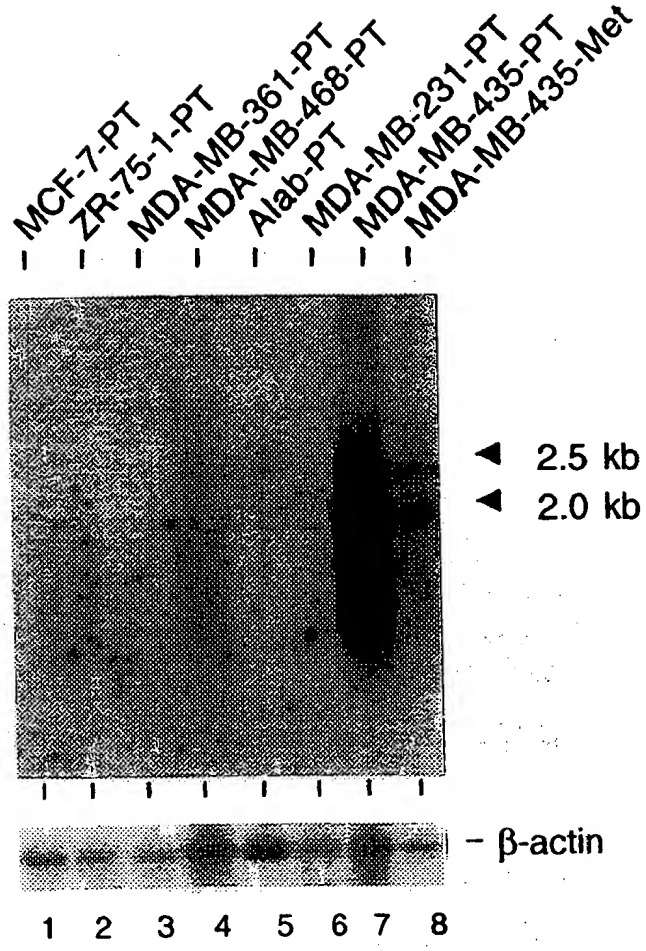


FIG. 4

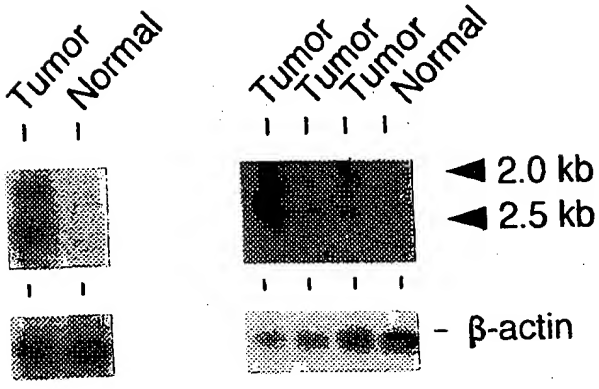


FIG. 5A FIG. 5B

H&E antisense sense

breast normal

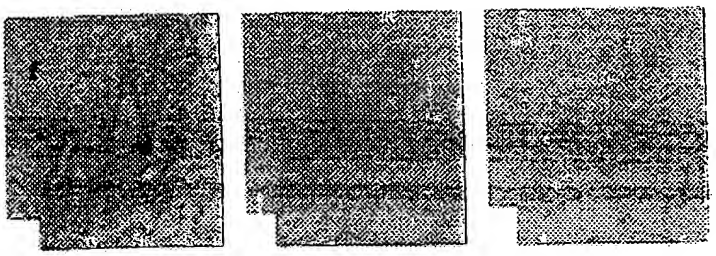


FIG. 6A FIG. 6B FIG. 6C

breast PT

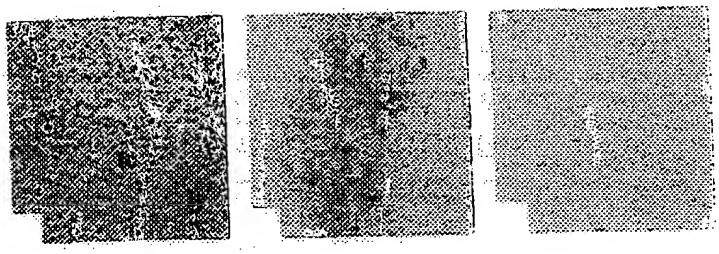


FIG. 6D FIG. 6E FIG. 6F

colon normal

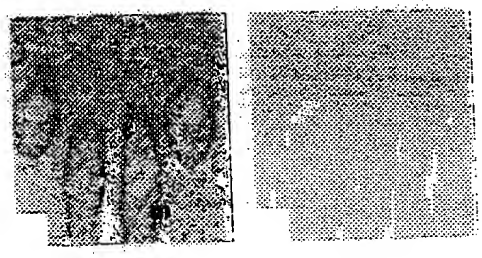


FIG. 6G FIG. 6H

colon PT

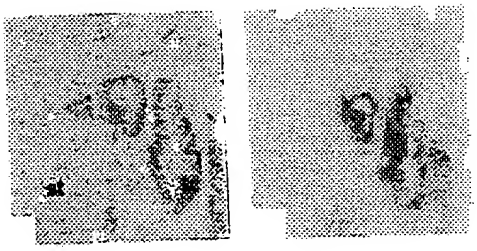


FIG. 6J FIG. 6K

liver met

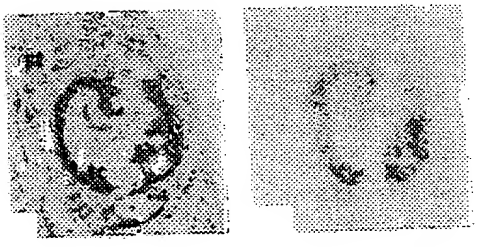


FIG. 6L FIG. 6M

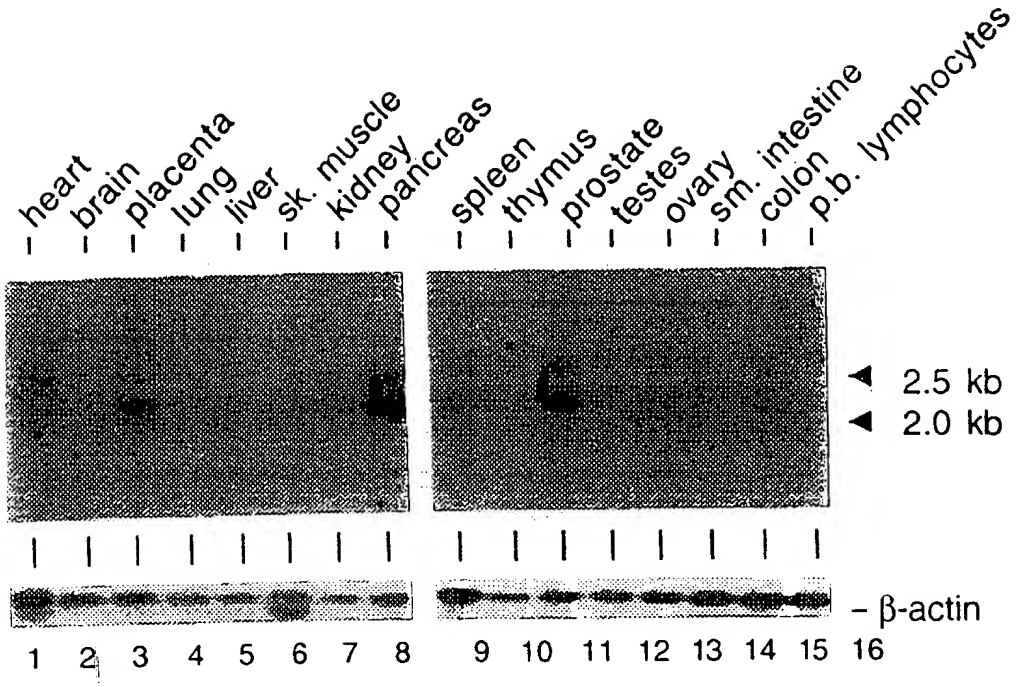

heart
brain
placenta
lung
liver
sk. muscle
kidney
pancreas
spleen
thymus
prostate
testes
ovary
sm. intestine
colon
p.b. lymphocytes
2.5 kb
2.0 kb
– β-actin
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

FIG. 7

INTERNATIONAL SEARCH REPORT

national Application No
PCT/US 98/26547

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/64 C12Q1/68 C12N15/57

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL (EMEST) Embl, Heidelberg Accession Number AA630313, 31 October 1997 HILLIER L. ET AL. : "ac08f06.s1 Stratagene HelA cell s3 937216 Homo sapiens cDNA clone 855875 3'" XP002101163 see abstract ---	8,11,13, 17
X	DATABASE EMBL (EMEST) EMBL, Heidelberg Accession Number AA293614, 22 April 1997 HILLIER L. ET AL.: "zt26c01.r1 Soares ovary tumour NbHOT Homo sapiens cDNA clone 71420" XP002101164 see abstract --- -/--	8,11,13, 17

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 April 1999	07/06/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Luzzatto, E

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/26547

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL (EMEST) EMBL, Heidelberg Accession Number N46495, 18 February 1996 HILLIER L. ET AL.: "yy42d05.r1 Homo sapiens cDNA clone 273897 5'" XP002101277 see abstract ---	3
P,X	EP 0 848 062 A (SMITHKLINE BEECHAM) 17 June 1998 see page 2, line 25 - line 30 see page 5, line 9 - line 32 see page 9, line 29 - line 57; tables 1,2 -----	1-33

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/26547

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 848062 A	17-06-1998	JP 11069981 A	16-03-1999